MULTIMEDIA UNIVERSITY

STUDENT ID NO

MULTIMEDIA UNIVERSITY

FINAL EXAMINATION

TRIMESTER 2, 2016/2017

DPS5018 – INTRODUCTION TO PROBABILITY AND STATISTIC

(B201-227, A81-92, I5)

8 MARCH 2017
2.30 p.m – 4.30 p.m
(2 Hours)

INSTRUCTIONS TO STUDENT

1. This question paper consists of **6** pages excluding cover page and appendix.
2. Attempt **ALL** questions.
3. Please write all your answers in the answer booklet provided.
4. Formulas are provided in the appendix section.

Please answer ALL questions and show the necessary working steps.

Question 1

a) Fill in the blank with the suitable choice of answer given in the box below.

i) _____ is the field of study implies collecting, organizing, presenting, analyzing and interpreting numerical data in order to make decision.

ii) A(An) _____ of a sample or population is a specific subject or object about which the information is collected.

iii) _____is a characteristic under study.

iv) The value of the variable for an element is called a(an) _____.

Observation	Element	Statistics	Variable
Measurement	Experimentation	Data	Census

(4 marks)

b) For each of the following, identify whether the statement is descriptive or inferential statistics. (4 marks)

No	Statement	Type of Statistics
i)	The department of Social Welfare and Development (DSWD) statistics recorded that there were five child abuse cases in past five years.	
ii)	A politician would like to estimate his chances for winning in the upcoming election based on opinion polls.	
iii)	A housewife wants to determine the average weekly amount she spent on groceries in the past 3 months.	
iv)	A new milk formulation designed to improve the psychomotor development of infant was tested on randomly selected infants. Based on the results, it was concluded that the new milk formulation is effective in improving the psychomotor development of infants.	

c) Identify the type of sampling used in each statement below.

i) Suppose a farmer wishes to work out the average milk yield of each cow type in his herd which consists of Ayrshire, Friesian, Galloway and Jersey cows. He could divide up his herd into the four sub-groups and take samples from each sub-groups. (1 mark)

ii) A number is assigned to 100 people at which the number is drawn by a judge blindly from a basket. (1 mark)

[TOTAL 10 MARKS]

Continued...

Question 2

a) The stem and leave diagram below shows the average intake of vitamin weekly, in grams, of 32 volunteers in a medical research.

Stem	Leaf									
8	1	2								
9	2	3	6	9						
10	0	2	4	5	5	6				
11	0	0	3	4	5	5	5	5	7	8
12	2	3	3	6	6	7				
13	3	3								
14	0	7								

Key: 10|0 means 100

i) List down the five-number summary of the distribution. (6.5 marks)
ii) Determine the interquartile range of the distribution. (1 mark)
iii) Identify the lower limit and upper limit of distribution. (2 marks)
iv) Draw a box plot to represent the above data. (3 marks)
v) Refer to stem and leaf diagram above, describe shape of the distribution. (1 mark)

b) A random sample of 8 independent measurements, in Pascal second, of the viscosity of a light machine oil is taken. The values obtained are given below.

25.2	24.8	25.0	24.0	24.5	24.6	25.0	25.6

Find the sample mean and standard deviation of the viscosity. (6.5 marks)

[TOTAL 20 MARKS]

Continued...

Question 3

a) The following table recorded the number of students who are taking the particular subjects for different courses in a particular semester.

Courses		DBA	DIA	DBIS	DIT
Subject	Statistics	118	24	10	7
	Mathematics	99	41	24	3
	Finance	33	45	16	0

Refer to contingency table above, answer question i) to iii).

i) If a student is selected from this group of students, what is the probability that the student is DIA student and is taking Mathematics? (1 mark)

ii) Which are the two events being mutually exclusive? (1 mark)

iii) Given that the selected student is a DBA student, determine the probability of he/she is taking Statistics. (4 marks)

b) A survey had been carried out in a group of 80 working adults, it is found that 28 adults like to join outdoor activities, 56 of them like indoor activities and 13 adults like both activities.

i) Represent the above situation in a Venn diagram. (3 marks)

ii) Determine the probability that randomly chosen adult likes to join any of these two activities. (2 marks)

c) The result in a study shows that 70% of the secondary school students in a town area go to school by bus. If they go to school by bus, the possibility that they will be late is 10%. Due to traffic jam issue, they will tend to be late to school by 80% if they are using other transport.

i) Draw a tree diagram for this situation. (3 marks)

ii) Find the probability that a student will be late to school. (2 marks)

iii) Given that a student will not be late to school, what is the probability that he/she goes to school by bus? (4 marks)

[TOTAL 20 MARKS]

Continued...

Question 4

a) The following probability distribution recorded number of television which was sent back for warranty in a factory.

Number of television, x	2	4	5	10	11
$P(X = x)$	0.42	0.36	0.18	m	0.03

i) Show that the value of m is 0.01. (2 marks)

ii) Find the mean of this distribution. (2 marks)

b) In Multimedia University Malacca campus, 80% of the students stay in hostel, IXORA apartment. If 20 students from this university are selected at random,

i) find the probability that 15 students stay in hostel. (3 marks)

ii) find the average number of students stay in hostel. (1 mark)

c) The mass of dark chocolate bars retailed by Beryl's factory in Seri Kembangan, Selangor have a normal distribution of mean mass of 250g and a standard deviation of 9g. The quality control officer will reject the product if its mass is less than 244.5g or more than 254.5g.

Calculate the probability that the officer will accept the product if a dark chocolate bar is chosen at random. (6 marks)

d) A condominium security guard claimed that they clamped an average of 5 cars which park illegally in condominium parking compound everyday.

i) Calculate the probability that at least 3 cars will be clamped per day.(3 marks)

ii) Calculate the probability that less than 11 cars will be clamped in 2 days. (3 marks)

[TOTAL 20 MARKS]

Continued...

Question 5

a) The following table shows the hours for sunshine, x, during the 10 days in January 2017 and the number of ice creams, y, sold by a beach shop in Maldives.

Sunshine Hours, x	Number of Ice Cream Sold, y
5	225
8	266
0	119
11	420
6	231
2	184
9	363
10	352
3	197
12	431

i) Calculate the total hours for sunshine which has been observed, $\sum x$, the total number of ice cream sold, $\sum y$ and $\sum xy$. (3 marks)

ii) Find the sum of squares, SS_{xx}, SS_{xy} and SS_{yy} if given that $\sum x^2 = 584$ and $\sum y^2 = 879402$. (3 marks)

iii) Obtain the least square regression equation in form of $\hat{y} = a + bx$. (4 marks)

iv) Find the correlation of coefficient, r. Determine the relationship between the sunshine hours, x and the number of ice cream sold, y. (3 marks)

v) The ice cream shop owner asks you to use the regression equation to predict the daily sales if there were 20 hours of sunshine. (2 marks)

b) In a study of the performance of MMU students in Statistics subject for the final examination, one of the factors influencing is the number of study hours in a week before the examination day. A study has been conducted on 40 boys and 40 girls. It has been claimed that the mean number of study hours for girls is always higher than the mean number of study hours for boys. The statistics are summarized below.

	Girls	Boys
Mean	4.2	3.6
Standard deviation	1.2310	1.3359

i) State the null and alternative hypotheses for the above situation. (2 marks)

ii) Calculate the test statistic and p-value by using the hypothesis testing for two population mean. (3 marks)

iii) At 5% significance level, can we conclude that the claim is true? (2 marks)

Continued...

c) A nutritionist recently proposed a new diet plan called sugar-restricted diet to diabetic patients. The nutritionist claimed that at least 20% of diabetic patients are following this new plan. Based on a random sample of 160 diabetic patients, 30 were on sugar-restricted diets. Test whether the claim is true at 0.01 significance level. (8 marks)

[TOTAL 30 MARKS]

End of Page.

APPENDIX – KEY FORMULA

- **Mean for ungrouped data**

$$\bar{x} = \frac{\sum x_i}{n} \quad \text{or} \quad \mu = \frac{\sum x_i}{N}$$

- **Mean for grouped data**

$$\bar{x} = \frac{\sum x_i f_i}{n} \quad \text{or} \quad \mu = \frac{\sum x_i f_i}{N}$$

Where x is the midpoint and f is the frequency of the class

- **Variance for ungrouped data**

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} \quad \text{or} \quad s^2 = \frac{\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}}{n-1}$$

- **Variance for grouped data**

$$s^2 = \frac{\sum (x_i - \bar{x})^2 f_i}{n-1} \quad \text{or} \quad s^2 = \frac{\sum x_i^2 f_i - \frac{\left(\sum x_i f_i\right)^2}{n}}{n-1}$$

Where x is the midpoint and f is the frequency of the class

DISCRETE RANDOM VARIABLE, BINOMIAL DISTRIBUTION, AND POISSON DISTRIBUTION

- Mean of a discrete random variable x : $\mu = \sum xP(X = x)$
- Standard deviation of a discrete random variable x : $\sigma = \sqrt{\sum x^2 P(X = x) - \mu^2}$
- Binomial probability formula : $P(X = x) = \binom{n}{x} p^x q^{n-x}$
- Poisson probability formula : $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$

- Key formula for Binomial Probability Distribution:
  - If $p \leq 0.5$
    i. $P(X = r) = B(r) - B(r-1)$
    ii. $P(X \geq r) = 1 - B(r-1)$
    iii. $P(X \leq r) = B(r)$
    iv. $P(X > r) = 1 - B(r)$
    v. $P(X < r) = B(r-1)$
    vi. $P(a \leq X \leq b) = B(b) - B(a-1)$

  - If $p > 0.5$
    i. $P(X = r) = B(n-r) - B(n-r-1)$
    ii. $P(X \geq r) = B(n-r)$
    iii. $P(X \leq r) = 1 - B(n-r-1)$
    iv. $P(a \leq X \leq b) = B(n-b) - B(n-a-1)$

- Key formula for Poisson Probability Distribution:
  - If $\mu = \lambda$
    i. $P(X = r) = Poi(r) - Poi(r-1)$
    ii. $P(X \geq r) = 1 - Poi(r-1)$
    iii. $P(X \leq r) = Poi(r)$
    iv. $P(X > r) = 1 - Poi(r)$
    v. $P(X < r) = Poi(r-1)$
    vi. $P(a \leq X \leq b) = Poi(b) - Poi(a-1)$

NORMAL AND STANDARD NORMAL PROBABILITY DISTRIBUTION

- z-value (observed value) for an x value : $Z = \dfrac{x - \mu}{\sigma}$

REGRESSION

- For the **least square regression**, $\hat{y} = a + bx$

  where $b = \dfrac{SS_{xy}}{SS_{xx}}$ and $a = \bar{y} - b\bar{x}$

- **Correlation coefficient**, r

  $$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}}$$

  where

  $$SS_{xx} = \sum x^2 - \frac{\left(\sum x\right)^2}{n}, \qquad SS_{yy} = \sum y^2 - \frac{\left(\sum y\right)^2}{n} \text{ and } \qquad SS_{xy} = \sum xy - \frac{\left(\sum x\right)\left(\sum y\right)}{n}$$

HYPOTHESIS TESTING

- **ONE population**

  Mean:

  $$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

  Proportion:

  $$Z = \frac{\hat{p} - p}{\sqrt{\dfrac{pq}{n}}}$$

- **TWO population**

  Mean:

  $$Z = \frac{\left(\bar{x}_1 - \bar{x}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\dfrac{\sigma_1^{\,2}}{n_1} + \dfrac{\sigma_2^{\,2}}{n_2}}}$$

  Proportion:

  $$Z = \frac{\left(\hat{p}_1 - \hat{p}_2\right) - \left(p_1 - p_2\right)}{\sqrt{p_p\left(1 - p_p\right)}\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} \qquad \textbf{where} \qquad p_p = \frac{x_1 + x_2}{n_1 + n_2}$$